

## SYSTEM AND METHOD FOR ANALYZING COMMUNICATIONS

This non-provisional application claims priority from U.S. Provisional Application No 60/168,242 filed December 1, 1999 and titled "ContactMap: Designing Our Digital Lives" the entire disclosure of which is incorporated by reference herein in its entirety as if fully reproduced herein.

### BACKGROUND OF THE INVENTION

The present invention relates generally to communications, and more particularly, to a system and a method for analyzing a user's collection of communications, such as electronic mail messages, telephone messages, voicemails, instant messaging dialogues and telephone logs to determine the identity of contacts within the communications, the relative priority of the contacts within the communications and any relationships between the contacts themselves and between the contacts and the user.

The ever-growing variety of mass communication devices and services such as wireless telephones, pagers, portable computers, personal digital assistants (PDAs), set top boxes, electronic mail, and instant messaging among others, has exponentially increased the number of different types of communications that one may receive on a daily basis. As a result, it has become very difficult to track one's personal and professional contacts since after even a short period of time, one could have hundreds of unique communications, such as electronic mail messages and voicemails to sort through.

Accordingly, it would be desirable to be able to easily analyze a collection of communications, such as electronic mail messages, to determine the identity of one's personal and professional contact, the relative priorities of these contacts and any existing relationship between and among these contacts.

## SUMMARY OF THE INVENTION

The invention provides a method for processing user communications to determine identities of contacts within the communications, to prioritize the relative importance of the contacts identified and to establish any relationships between the contacts themselves and between the contacts and the user. In one embodiment, the method includes the steps of receiving a plurality of communication files, wherein the communication files include a plurality of different contacts, reviewing the plurality of communication files to determine contact identities, contact priorities and relationships and providing an analysis of the communication files, wherein the analysis provides the identities of contacts identified in the communication files, the relative importance of contacts relative to one another and any relationships between two or more contacts.

In the present invention, a contact may be refer to an individual person, a company, a group of individual people, an organization, an electronic mail listing or any combination or variations of the aforementioned. Contacts may be related to a user's personal life, workplace or other social networks. Typically, the user will specify a directory, folder or other repository which contains the communications, such as the electronic mail messages to be analyzed. A communications or message processor will receive the communications and from these communications, determine at least the identity of the contacts and their relative priority to one another to the user. For example, contacts which engage in frequent, reciprocal communications with the user are rated or assigned a higher priority than contacts which only engage in one-way communications with the user. Though the sheer number of communications a contact may have with the user is relevant in determining the contact priority, the reciprocal nature of the communications is more indicative as to the importance of the communications.

In one embodiment of the present invention, contacts which are frequently mentioned or "co-mentioned" within communication may also be classified as being related to a similar grouping or network in relation to the user. For example, co-workers who frequently copy each other on electronic mail message may be considered to be part of a single grouping of contacts.

## BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates an exemplary system in accordance with the teachings of the present invention.

5

FIG. 2 illustrates an exemplary method in accordance with the teachings of the present invention.

10

FIG. 3 illustrates another exemplary method in accordance with the teachings of the present invention.

15

FIG. 4 illustrates yet another exemplary method in accordance with the teachings of the present invention.

20

FIG. 5 illustrates still another exemplary method in accordance with the teachings of the present invention.

FIG. 6 illustrates an exemplary message processing display in accordance with the teachings of the present invention.

FIGS. 7a-7d illustrate an exemplary code implementation of a sorting method in accordance with the teachings of the present invention.

FIG. 8a-8b illustrate an exemplary code implementation of a clustering method in accordance with the teachings of the present invention.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a system and method for analyzing a user's communications, such as electronic mail messages, voicemail messages, instant messages, telephone logs and other such communications. The communications are analyzed to determine the identity of any contacts within the communications, the priority of the contacts within the communications and any inferred relationships between the contacts themselves and between the contact(s) and the user. As used herein, the term contact is used to refer to an individual person, a company, a group of individual people, an organization, an electronic mail listing or a combination of the aforementioned with which a user may communicate and/or receive communications from. For example, a user's colleagues in a work environment may be considered contacts as well as members of a user's family may be considered contacts. Contacts may also be part of and classified into contact groupings, such as related contacts in a family grouping or colleagues in a work grouping.

Referring to FIG. 1, a system 10 of the present invention includes a communications processor 20 which analyzes a user's communications, such as the user's electronic mail messages 30, voicemail messages 40, instant messages 50, telephone logs 60 and/or other communications 70 to determine contact identities, priorities and relationships. In a preferred embodiment, the communications processor 20 may be configured as specialized software which implements the methods disclosed here and may be run on a computing device, such as a personal computer, personal digital assistant, wireless phone or other similar device. Such specialized software which implements the methods described herein may also be used in a network device, such as a network server or electronic message server, such as an electronic mail message server. In the present invention, the communications processor 20 will receive and analyze the communications, such as electronic mail messages 20, and provide a contact network display 80 to the user. In one embodiment, the contact display 80 may include a listing of the identified contacts from the communications. Such a listing may be prioritized and arranged in a manner according to the rankings of the contacts as determined from an analysis of the communications, as shown and described later herein.

Referring to FIG. 2, a first generalized embodiment of a process for analyzing a user's communications in accordance with the teachings of the present invention is shown. In this embodiment, a plurality of communications are accessed or extracted, step 200, such as by communications processor 20 previously shown in FIG. 1. Once accessed, the communications are processed to determine the relative importance of contacts based on the communications, step 210, such as by the communications processor 20. Once processed, a contact display 220 is provided to the user, step 220, which may include clusters of related contacts. The user is then free to approve and/or modify the contact display as the user sees fit.

Referring to FIG. 3, another embodiment for analyzing a user's communication is shown. As shown in FIG. 3, a collection of communication messages, such as electronic mail messages, are received from a user, step 300. In one embodiment of the present invention, the user may specify certain folders, directories, networks or other locations in which these communication messages are stored. These communication messages are processed, step 310, which may include determining the amount of messages from certain specific contacts, an analysis of the message content and/or a determination if the user and the contact have participated in mutual reciprocal communications. Once the messages are processed, a contact ranking is determined based on the processing of the messages, step 320. A contact display of the user's identified contacts is then provided to the user, step 330. In alternative embodiments, other information may be provided to the user via the contact display, such as the priority ratings of each contact and any determined relationships between contacts and the user.

In one embodiment of the present invention, the communications analysis process may be accomplished by initially having a user provide or specify one or more electronic mail folders to process. Once the electronic mail folders are specified, electronic mail message header information from the electronic mail messages are analyzed to determine a relative importance of entities or contacts mentioned in the messages. For electronic mail messages, the header fields, such as the to:, from:, cc:, and bcc: fields may be analyzed to determine which contacts are mentioned or referenced in the messages. In one embodiment, greater priority/weight is attached to contacts who engage in frequent, reciprocal communication (i.e., the user responds to

messages from the contact and vice versa), as opposed to if the user sends or receives communications without any reciprocity.

Referring to FIG. 4, another embodiment of the present method for analyzing communications is shown. In this embodiment, a user may be prompted to identify a collection of existing communication files, such as a directory or folder of electronic mail messages, step 400. The identified communication files are accessed, step 410, for example, by accessing a network directory or folder containing the electronic mail messages. Once accessed, the communication files are processed, step 420. Processing the communication files, step 420, may include determining the identities of contacts within the communication files, substep 422; determining the priorities of contacts relative to one another and/or relative to the user, substep 424; and determining any relationships among the contacts and/or the user, substep 426. Once the communication files have been processed, the user's contact preferences are received, step 430. The user's contact preferences may be in form of approvals and disapprovals of the contact identities, priorities and arrangements determined in step 420. For example, if a contact X is determined to be a higher priority contact than a contact Y, the user may override such a determination and assign a higher priority to contact Y even though contact X was previously determined to be of higher priority.

In the present invention, it is contemplated that users can process their communication files, such as their electronic mail messages, via the communications analysis process described herein at any time to create a contact network or incrementally update their contact network display. For example, the message or communications processing may be used to identify new contacts within a user's electronic mail messages and possibly add these new contacts to the user's workspace. Such an updating is best performed at some incremental time interval to ensure that the user's contact network contains the most updated set of contacts.

Referring to FIG. 5, another embodiment of the present method for analyzing a user's communications is shown. In this embodiment, one or more contacts are identified within a communication, step 500, such as by reviewing a header portion of an electronic mail message and/or voicemail message. A determination is made whether the identified contact or contacts

have engaged in frequent communications with the user, step 510. Such a determination may be made, for example, by simply counting the number of messages a contact either sent to the user and/or received from the user. If the contact or contacts have engaged in frequent communications with the user, the contact or contacts are assigned a higher priority, step 520. If the contact or contacts have not engaged in frequent communications with the user, the contact or contacts are assigned a relative lower priority, step 530. The threshold for what constitutes frequent communications may be set at any level. For example, the threshold may be set at ten (10) communications, such that if it is determined that a contact engaged in at least ten (10) communications with the user, that contact will then be assigned a higher priority.

Referring again to FIG. 5, another determination is made to decide whether the communications between the user and the contact or contacts are reciprocal communications, e.g., if the communication from the contact is in response to an initial communication from the user, step 540. Reciprocal communications can also be initiated by a contact and thus responses by the user to an initial communication by the contact may also be considered to be reciprocal communications. If the contact or contacts have engaged in reciprocal communications with the user, the contact or contacts are assigned a higher priority, step 550. If the contact or contacts have not engaged in reciprocal communications with the user, the contact or contacts are assigned a relative lower priority, step 560. For example, if a user receives a message from a particular contact, and the user does not reply to the message, the contact may be given a lower priority rating. Additionally, if the user continually receives message from a particular contact and never replies to the message, the contact may be given an even lower priority rating.

Referring again to FIG. 5, a determination may be made based on the recency of the communication, step 570, e.g. how old the communication is relative to other communications being analyzed. Such a determination may be made, for example, by simply processing the date field in an electronic mail message header. If it is determined that the communication is a recent communication, the contact associated with the recent communication may be assigned a higher priority, step 580. If the communication is determined to be an old or a relatively older communication, the contact associated with the communication may be assigned a relative lower priority, step 590. The threshold for what constitutes a recent communication may be set at any

level. For example, the threshold may be configured such that any communication that is less than a month old can be considered recent and any communication received that is older than a month may be considered not recent. Furthermore, additional temporal information such as the recency and duration of interactions or rapidity of responses in reciprocal interactions may be analyzed in determining the priority of contacts for a user. For example, contacts which engage in relatively longer duration communications and/or correspond with a user in rapid, reciprocal communications may be rated higher than contacts which do not. For example, if a contact often replies to a message from the user, or the user frequently replies to messages from that particular contact, then it may be inferred that the contact is important to the user. Another strong criterion for inferring importance is if the user often sends messages to the contact. Frequent attempts by the user to target the contact may be some indication of the contact's value. Weaker criteria are that the Contact often sends messages to the user, or is mentioned in (but does not originate) messages to and from the user.

Referring yet again to FIG. 5, another determination is made to decide whether there are any co-mentions in the communications, step 592. As used herein, the term "co-mentions" refers to certain contacts that are referenced or mentioned within the user's communications, such as in their electronic mail messages. For messages or communications involving more than a sender and a single recipient, the group or set of contacts that are mentioned together in the same message will be referenced as "co-mentions". Relations are thus inferred between contacts on the basis of these co-mentions. For example, the more that people are mentioned together, the stronger the inferred relationship, as discussed in more detail later herein.

In one embodiment, the system message processing or communications analysis determines whether or not contacts are mentioned in the header fields of the same electronic mail message. For example, if two people are repeatedly mentioned in the "to:" field of an electronic mail message header from a given contact, then an inference may be made that there is a relationship between that contact and the two co-mentioned people. This information about relationships has two potential benefits: (a) it allows the construction of groups, with straightforward addressing since instead of having to set up electronic mail aliases for groups a user repeatedly communicates with, these groups may be detected automatically and can be used



for rapid addressing of messages to multiple individuals; (b) co-mentions may be useful in guiding contact layout, e.g. with co-mentioned contacts being near one another in the initial automatic layout configuration of the user's workspace display. In another embodiment of the present invention, co-mentions may be determined by determining whether a contact routinely forwards a message from another contact to the user or vice versa. In such a case, the contact from which the forwarded message originated, may be considered to be related to the contact which forwarded the message and/or also the user. Referring back to FIG. 5, if two or more contacts have been co-mentioned within the communications, a relationship between the contacts themselves and/or the user is inferred, step 594. If there are no co-mentioning of contacts within the communications, no relationship between the contacts themselves and/or the user is inferred, step 598.

In one embodiment of the present invention, importance ranking of contact(s) may be represented as a linear equation using the following weightings on the above criteria:

$$\text{Importance score} = Kx_1 + Lx_2 + Mx_3 + Nx_4,$$

where

- $x_1$  = number of messages the user replies to from the originating contact;
- $x_2$  = number of messages the user sends to the contact excluding replies;
- $x_3$  = number of reply messages from the contact;
- $x_4$  = combined total of messages that are not composed by the user in which the contact is mentioned in the cc or to lines, plus messages from the contact that are not replies;
- K, L, M and N are constant weightings. K applies to messages involving reciprocity and is very large to respect the significance of reciprocal interactions initiated by the user. L is moderately large, based on the intuition that if the user initiates a communication with a contact this suggests that they attach importance to that person. M is also moderately large to respect the importance of contact initiated interactivity. N is small based on the proposition that little value is attached to messages initiated by other people that are not responses, e.g. some of which may be "spam". Similarly, passing mentions of a contact name in messages, to, and from, others may not be highly rated.

Referring now to FIG. 6, an exemplary contact network display or listing 600 is shown. After the communications analysis or processing as described above is performed, the user may be provided with such a display or listing 600. In one embodiment, the display 600 includes a name section 610, an address section 620, an oldest message date section 630 and a newest message date section 640 and a score section 650. An approval or "include" section 660 may also be provided to the user, which allows the user to include or exclude any number of contacts from the user's contact network.

In the present invention, any number of scoring schemes may be used to rate or prioritize contacts based on the communications processing discussed herein. For example, a 1 to 300 scale may be used, a 1 to 10 scale may also be used as well as an A-Z rating scheme. Rating points or allocations may be provided to a contact in also a number of different manners. For example, a contact may be assigned or awarded a single point for every piece of communication, i.e. electronic mail message, voicemail message, instant message, etc. that the contact participated in with the user. For example, if a certain contact sent ten electronic mail messages to the user, the contact will be awarded 10 point under this exemplary scoring scheme. The contact's cumulative score may be adjusted by a number of factors, such as by increasing the score if the contact and the user engaged in reciprocal rather than one-way communications. The term "one-way" communications being defined as if the contact sent the user certain communications, but the user never responded to the communications, such as may be the case in terms of "SPAM" or unsolicited electronic mail. Using another exemplary rating or scoring scheme, the contact score may be modified based on certain portions of the communication's content, for example, if the content of a message is determined to have many informal, slang, or otherwise colloquial uses of language in the message content, the contact may be considered to be a personal rather than a business or professional contact and then rated higher or lower accordingly depending on the preferences of the user.

In implementing the teachings of the present invention, it is preferable to develop the system and method herein with an object-oriented structure which may be written in an object-oriented language such as Java. One exemplary implementation may include a set of higher-

level user interface components on top of an underlying persistent contact-based data layer and message analysis component. Data objects may be strictly separated from their display to allow multiple display components to display the same underlying data. The Java delegation event model may be used to allow user interface components to easily track changes to underlying data objects. Java object serialization may also be used to persistently store contact information. Referring to FIGS. 7a-7d, an exemplary code listing is shown which may be used to prioritize contacts as discussed herein is shown. Referring to FIGS. 8a-8b, an exemplary code listing is shown which may be used to cluster contact in accordance with the teachings herein.

10

It will be apparent to those skilled in the art that many changes and substitutions can be made to the system and method described herein without departing from the spirit and scope of the invention as defined by the appended claims. For example, natural language analysis techniques and programs that analyze, for example, electronic mail messages, may be used. Additionally, voice recognition analysis techniques and programs may also be used to analyze the communications, such as voicemail messages. For example, voice recognition analysis may be used to convert voicemail messages into text. The text may then be processed to extract certain contact information from the text, such as contact names, numbers, etc.

15